



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 11:04 PM GMT

PDB ID : 5FPT  
Title : Structure of hepatitis C virus (HCV) full-length NS3 complex with small-molecule ligand 2-(1-methyl-1H-indol-3-yl)acetic acid (AT3437) in an alternate binding site.  
Authors : Jhoti, H.; Ludlow, R.F.; Saini, H.K.; Tickle, I.J.; Verdonk, M.; Pathuri, P.; Williams, P.A.  
Deposited on : 2015-12-02  
Resolution : 2.72 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026982
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20026982

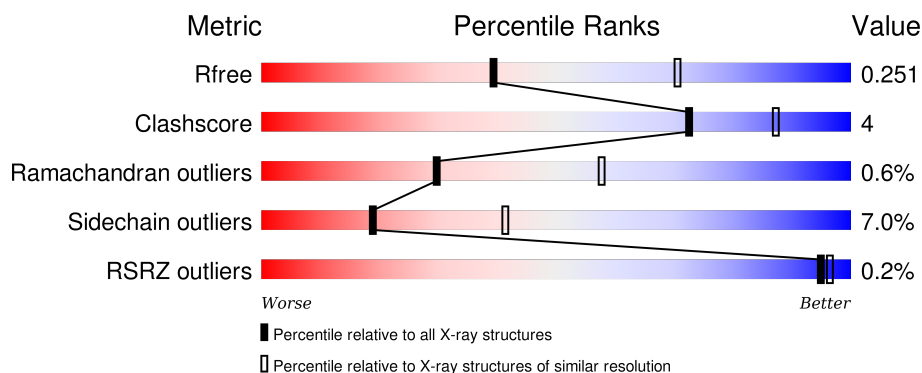
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*


The reported resolution of this entry is 2.72 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	2439 (2.74-2.70)
Clashscore	102246	2771 (2.74-2.70)
Ramachandran outliers	100387	2726 (2.74-2.70)
Sidechain outliers	100360	2727 (2.74-2.70)
RSRZ outliers	91569	2443 (2.74-2.70)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	666	 82% 13% . .
1	B	666	 82% 13% . .

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9911 atoms, of which 20 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called HEPATITIS C VIRUS FULL-LENGTH NS3 COMPLEX.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	645	Total	C	N	O	S	0	0	0
			4701	2974	791	906	30			
1	B	642	Total	C	N	O	S	0	0	0
			4699	2969	803	897	30			

There are 82 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	684	MET	-	EXPRESSION TAG	UNP P26663
A	685	GLY	-	EXPRESSION TAG	UNP P26663
A	686	SER	-	EXPRESSION TAG	UNP P26663
A	687	SER	-	EXPRESSION TAG	UNP P26663
A	688	HIS	-	EXPRESSION TAG	UNP P26663
A	689	HIS	-	EXPRESSION TAG	UNP P26663
A	690	HIS	-	EXPRESSION TAG	UNP P26663
A	691	HIS	-	EXPRESSION TAG	UNP P26663
A	692	HIS	-	EXPRESSION TAG	UNP P26663
A	693	HIS	-	EXPRESSION TAG	UNP P26663
A	694	SER	-	EXPRESSION TAG	UNP P26663
A	695	SER	-	EXPRESSION TAG	UNP P26663
A	696	GLY	-	EXPRESSION TAG	UNP P26663
A	697	LEU	-	EXPRESSION TAG	UNP P26663
A	698	VAL	-	EXPRESSION TAG	UNP P26663
A	699	PRO	-	EXPRESSION TAG	UNP P26663
A	700	ARG	-	EXPRESSION TAG	UNP P26663
A	701	GLY	-	EXPRESSION TAG	UNP P26663
A	702	SER	-	EXPRESSION TAG	UNP P26663
A	703	HIS	-	EXPRESSION TAG	UNP P26663
A	704	MET	-	EXPRESSION TAG	UNP P26663
A	705	GLY	-	EXPRESSION TAG	UNP P26663
A	706	SER	-	EXPRESSION TAG	UNP P26663
A	707	VAL	-	EXPRESSION TAG	UNP P26663
A	708	VAL	-	EXPRESSION TAG	UNP P26663

*Continued on next page...*

*Continued from previous page...*

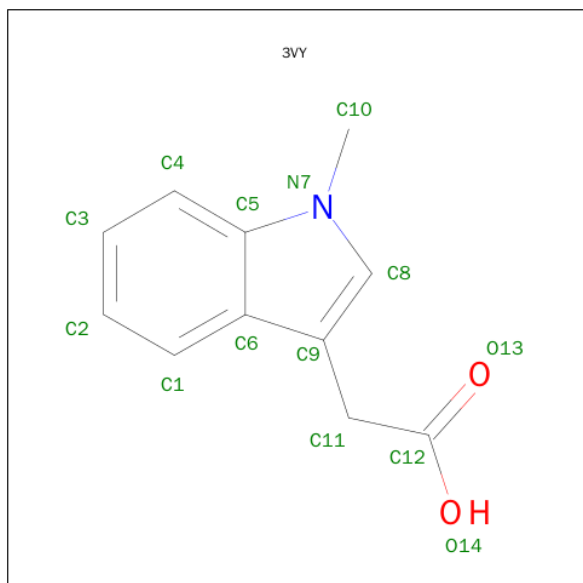
Chain	Residue	Modelled	Actual	Comment	Reference
A	709	ILE	-	EXPRESSION TAG	UNP P26663
A	710	VAL	-	EXPRESSION TAG	UNP P26663
A	711	GLY	-	EXPRESSION TAG	UNP P26663
A	712	ARG	-	EXPRESSION TAG	UNP P26663
A	713	ILE	-	EXPRESSION TAG	UNP P26663
A	714	ILE	-	EXPRESSION TAG	UNP P26663
A	715	LEU	-	EXPRESSION TAG	UNP P26663
A	716	SER	-	EXPRESSION TAG	UNP P26663
A	717	GLY	-	EXPRESSION TAG	UNP P26663
A	718	SER	-	EXPRESSION TAG	UNP P26663
A	719	GLY	-	EXPRESSION TAG	UNP P26663
A	720	SER	-	EXPRESSION TAG	UNP P26663
A	66	GLY	ALA	CONFLICT	UNP P26663
A	86	GLN	PRO	CONFLICT	UNP P26663
A	87	ALA	LYS	CONFLICT	UNP P26663
A	147	SER	PHE	CONFLICT	UNP P26663
B	684	MET	-	EXPRESSION TAG	UNP P26663
B	685	GLY	-	EXPRESSION TAG	UNP P26663
B	686	SER	-	EXPRESSION TAG	UNP P26663
B	687	SER	-	EXPRESSION TAG	UNP P26663
B	688	HIS	-	EXPRESSION TAG	UNP P26663
B	689	HIS	-	EXPRESSION TAG	UNP P26663
B	690	HIS	-	EXPRESSION TAG	UNP P26663
B	691	HIS	-	EXPRESSION TAG	UNP P26663
B	692	HIS	-	EXPRESSION TAG	UNP P26663
B	693	HIS	-	EXPRESSION TAG	UNP P26663
B	694	SER	-	EXPRESSION TAG	UNP P26663
B	695	SER	-	EXPRESSION TAG	UNP P26663
B	696	GLY	-	EXPRESSION TAG	UNP P26663
B	697	LEU	-	EXPRESSION TAG	UNP P26663
B	698	VAL	-	EXPRESSION TAG	UNP P26663
B	699	PRO	-	EXPRESSION TAG	UNP P26663
B	700	ARG	-	EXPRESSION TAG	UNP P26663
B	701	GLY	-	EXPRESSION TAG	UNP P26663
B	702	SER	-	EXPRESSION TAG	UNP P26663
B	703	HIS	-	EXPRESSION TAG	UNP P26663
B	704	MET	-	EXPRESSION TAG	UNP P26663
B	705	GLY	-	EXPRESSION TAG	UNP P26663
B	706	SER	-	EXPRESSION TAG	UNP P26663
B	707	VAL	-	EXPRESSION TAG	UNP P26663
B	708	VAL	-	EXPRESSION TAG	UNP P26663
B	709	ILE	-	EXPRESSION TAG	UNP P26663

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	710	VAL	-	EXPRESSION TAG	UNP P26663
B	711	GLY	-	EXPRESSION TAG	UNP P26663
B	712	ARG	-	EXPRESSION TAG	UNP P26663
B	713	ILE	-	EXPRESSION TAG	UNP P26663
B	714	ILE	-	EXPRESSION TAG	UNP P26663
B	715	LEU	-	EXPRESSION TAG	UNP P26663
B	716	SER	-	EXPRESSION TAG	UNP P26663
B	717	GLY	-	EXPRESSION TAG	UNP P26663
B	718	SER	-	EXPRESSION TAG	UNP P26663
B	719	GLY	-	EXPRESSION TAG	UNP P26663
B	720	SER	-	EXPRESSION TAG	UNP P26663
B	66	GLY	ALA	CONFLICT	UNP P26663
B	86	GLN	PRO	CONFLICT	UNP P26663
B	87	ALA	LYS	CONFLICT	UNP P26663
B	147	SER	PHE	CONFLICT	UNP P26663

- Molecule 2 is (1-METHYL-1H-INDOL-3-YL)ACETIC ACID (three-letter code: 3VY) (formula: C<sub>11</sub>H<sub>11</sub>NO<sub>2</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	0	0
			24	11	10	1	2		
2	B	1	Total	C	H	N	O	0	0
			24	11	10	1	2		

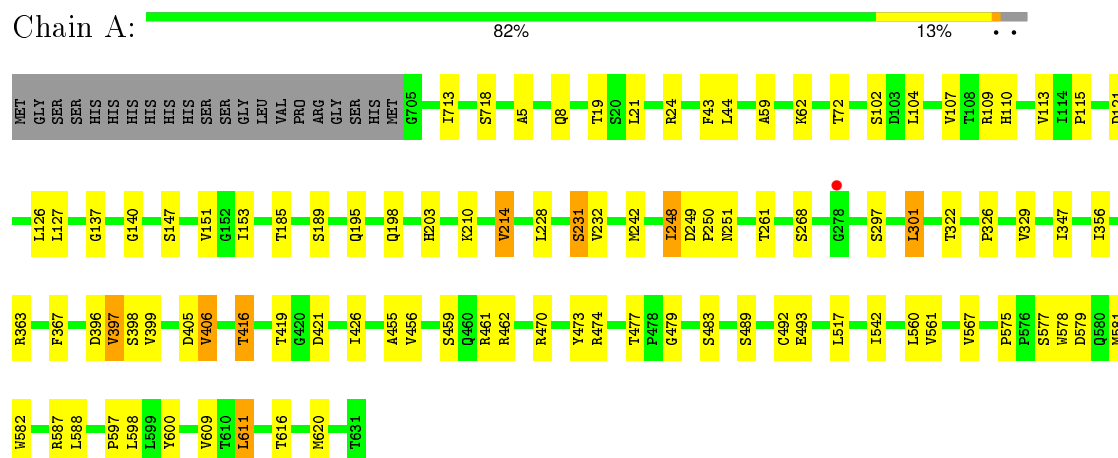
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	282	Total 282	O 282	0	0
3	B	181	Total 181	O 181	0	0

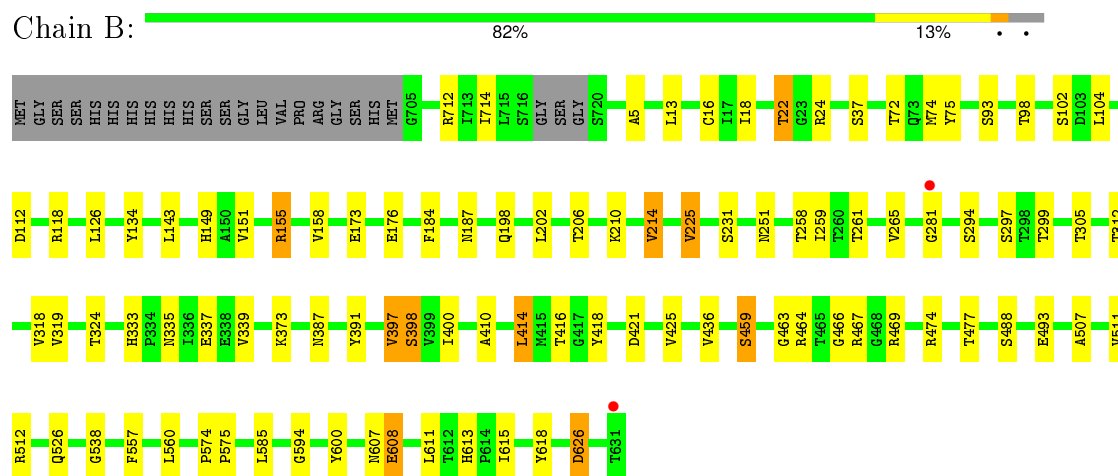
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

#### • Molecule 1: HEPATITIS C VIRUS FULL-LENGTH NS3 COMPLEX



#### • Molecule 1: HEPATITIS C VIRUS FULL-LENGTH NS3 COMPLEX



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	91.60Å 110.62Å 140.78Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	70.55 – 2.72 70.55 – 2.72	Depositor EDS
% Data completeness (in resolution range)	98.4 (70.55-2.72) 98.4 (70.55-2.72)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.99 (at 2.73Å)	Xtriage
Refinement program	BUSTER 2.11.6	Depositor
R, $R_{free}$	0.155 , 0.251 0.161 , 0.251	Depositor DCC
$R_{free}$ test set	1955 reflections (5.35%)	DCC
Wilson B-factor (Å <sup>2</sup> )	45.8	Xtriage
Anisotropy	0.387	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 76.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 38697 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	9911	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.18% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.



## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 3VY

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	0.64	0/4810	0.76	0/6587
1	B	0.64	0/4805	0.75	0/6569
All	All	0.64	0/9615	0.75	0/13156

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4701	0	4537	39	0
1	B	4699	0	4535	39	0
2	A	14	10	10	0	0
2	B	14	10	10	0	0
3	A	282	0	0	2	0
3	B	181	0	0	1	0
All	All	9891	20	9092	77	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 4.

The worst 5 of 77 close contacts within the same asymmetric unit are listed below, sorted by their

clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:74:MET:HG2	1:B:75:TYR:CE2	2.19	0.77
1:B:613:HIS:HD2	1:B:615:ILE:H	1.37	0.70
1:B:18:ILE:O	1:B:22:THR:HG23	1.93	0.69
1:B:294:SER:HB2	1:B:299:THR:HG21	1.76	0.68
1:A:611:LEU:HD23	1:B:611:LEU:HD23	1.77	0.66

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	643/666 (96%)	614 (96%)	25 (4%)	4 (1%)	30	58
1	B	636/666 (96%)	601 (94%)	31 (5%)	4 (1%)	30	58
All	All	1279/1332 (96%)	1215 (95%)	56 (4%)	8 (1%)	30	58

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	398	SER
1	A	398	SER
1	A	542	ILE
1	B	463	GLY
1	B	387	ASN

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	490/544 (90%)	457 (93%)	33 (7%)	20	44
1	B	484/544 (89%)	449 (93%)	35 (7%)	18	40
All	All	974/1088 (90%)	906 (93%)	68 (7%)	19	41

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	517	LEU
1	B	37	SER
1	B	416	THR
1	A	600	TYR
1	B	13	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	77	ASN
1	B	201	HIS
1	B	335	ASN
1	B	27	ASN
1	B	333	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	3VY	A	1721	-	9,15,15	0.72	0	11,21,21	0.49	0
2	3VY	B	1721	-	9,15,15	0.78	0	11,21,21	0.51	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	3VY	A	1721	-	-	0/2/4/4	0/2/2/2
2	3VY	B	1721	-	-	0/2/4/4	0/2/2/2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	645/666 (96%)	-0.31	1 (0%) 95 96	23, 43, 68, 88	0
1	B	642/666 (96%)	-0.21	2 (0%) 94 95	28, 47, 71, 100	0
All	All	1287/1332 (96%)	-0.26	3 (0%) 95 96	23, 46, 69, 100	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	281	GLY	2.5
1	A	278	GLY	2.1
1	B	631	THR	2.1

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	3VY	A	1721	14/14	0.97	0.17	1.35	14,24,26,26	24
2	3VY	B	1721	14/14	0.96	0.15	-0.02	55,60,66,66	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.